available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/agwat

Agricultural Water Management

# Model for predicting rainfall by fuzzy set theory using USDA scan data

**M. Hasan [a,*], T. Tsegaye [a,1], X. Shi [b,2], G. Schaefer [c,3], G. Taylor [d,4]**

[a] Dept. of Natural Resources and Environmental Sciences, Alabama Agricultural and Mechanical University (AAMU), 4900 Meridian St., Normal, AL 35762, USA

[b] Dept. of Computer Science, Alabama Agricultural and Mechanical University (AAMU), 4900 Meridian St., Normal, AL 35762, USA

[c] NRCS-National Water and Climate Center, 1201 NE Lloyd Blvd., Suite 802, Portland, OR 97232, USA

[d] Oregon Climate Service, 328 Strand Agricultural Hall, Oregon State University, Corvallis, OR 97331, USA

ARTICLE INFO

*Article history:*
Received 30 July 2007
Accepted 27 July 2008
Published on line 23 October 2008

*Keywords:*
Production rules
Regression coefficient
Fuzzy inference model
Fuzzy levels
Membership function

ABSTRACT

This paper presents a fuzzy inference model for predicting rainfall using scan data from the USDA Soil Climate Analysis Network Station at Alabama Agricultural and Mechanical University (AAMU) campus for the year 2004. The model further reflects how an expert would perceive weather conditions and apply this knowledge before inferring a rainfall. Fuzzy variables were selected based on judging patterns in individual monthly graphs for 2003 and 2004 and the influence of different variables that cause rainfall. A decrease in temperature (TP) and an increase in wind speed (WS) when compared between the ith and (i − 1)th day were found to have a positive relation with a rainfall (RF) occurrence in most cases. Therefore, TP and WS were used in the antecedent part of the production rules to predict rainfall (RF). Results of the model showed better performance when threshold values for: (1) relative humidity (RH) of ith day, (2) humidity increase (HI) between the ith and (i − 1)th day, and (3) product (P) of decrease in temperature (TP) and an increase in wind speed (WS) were introduced. The percentage of error was 12.35 when compared the calculated amount of rainfall with actual amount of rainfall.

Published by Elsevier B.V.

## 1. Introduction

In predicting weather conditions, factors in the antecedent and consequent parts that exhibit vagueness and ambiguity are being treated with logic and valid algorithms (Hasan et al., 1995). Use of fuzzy set theory has been proved by scientists to be applicable with uncertain, vague and qualitative expressions of the system. Application of fuzzy set theory in soil, crop, and water management is still in its infant stage due to the lack of awareness of the potentials of fuzzy set theory. Weather forecasting is one of the most important and demanding operational responsibilities carried out by meteorological services worldwide. It is a complicated procedure that includes numerous specialized technological fields. The task

* *Corresponding author*. Tel.: +1 256 372 4560; fax: +1 256 372 8404.
E-mail addresses: mahbub.hasan@aamu.edu (M. Hasan), teferi.tsegaye@aamu.edu (T. Tsegaye), frank.shi@aamu.edu (X. Shi), garry.schaefer@por.usda.gov (G. Schaefer), taylor@coas.oregonstate.edu (G. Taylor).

[1] Tel.: +1 256 372 4219; fax: +1 256 372 8404.

[2] Tel.: +1 256 372 4144; fax: +1 256 372 5578.

[3] Tel.: +1 503 414 3068; fax: +1 503 414 3101.

[4] Tel.: +1 541 737 5705; fax: +1 541 737 5710.

0378-3774/$ – see front matter. Published by Elsevier B.V.
doi:10.1016/j.agwat.2008.07.015

is complicated in the field of meteorology because all decisions are made within a visage of uncertainty associated with weather systems. Chaotic features associated with atmospheric phenomena have also attracted the attention of modern scientists. The drawback of statistical models is dependence, in most cases, upon several tacit assumptions regarding the system (Wilks, 1998). Carrano et al. (2004) compared non-linear regression modeling and fuzzy knowledge-based modeling, and explained that fuzzy models are most appropriate when subjective and qualitative data are utilized and the numbers of empirical observations are small. Brown-Brandl et al. (2003) used four modeling techniques to predict respiration rate as an indicator of stress in livestock. Four modeling techniques consisted of two multiple regression and two fuzzy inference systems. Fuzzy inference models offered better results than the two multiple regression models (Brown-Brandl et al., 2003). Fuzzy inference models yield a lower percentage of error when compared to the linear multiple regression model (Hasan et al., 1995). Similar research by Wong et al. (2003) compared the results of fuzzy rule-based rainfall prediction with an established method which uses radial basis function networks and orographic effect. They concluded that fuzzy rule-based methods could provide similar results from the established method. However, the method has an advantage of allowing the analyst to understand and interact with the model using fuzzy rules. Lee et al. (1998) considered two smaller areas where they assumed precipitation is proportional to elevation. Predictions of those two areas were made using a simple linear regression based on elevation information only. Comparison with the observed data revealed that the radial basis function (RBF) network produced better results than the linear regression models. Hence, considering the advantage of using the concept of fuzzy logic for predicting rainfall as stated by other researchers is justifiable. The advantage of fuzzy inference modeling can reflect expert knowledge and yield results with precision and accuracy. In fuzzy rule basics, knowledge acquisition is the main concern for building an expert system. Knowledge in the form of IF–THEN rules can be provided by experts or can be extracted from data. Each rule has an antecedent part and a consequent part. The antecedent part is the collection of conditions connected by AND, OR, NOT logic operators and the consequent part represents its action (Pant and Ashwagosh, 2004). In a fuzzy inference engine, the truth-value for the premise of each rule is computed and applied to the conclusion part of each rule. This result is one fuzzy subset being assigned to each output variable for each rule. For composite rules usually min–max inference technique is used.

Defuzzification is used to convert fuzzy output sets to a crisp value. The widely used methods for defuzzification are center of gravity and mean of maxima.

Generating production rules for fuzzy inference modeling is cumbersome if they are not derived as they are being perceived by an expert. Production rules have the form:

$$\text{IF } X \text{ is } A_1 \text{ AND } Y \text{ is } B_1 \text{ THEN } Z \text{ is } C_1 \tag{1}$$

$$\text{IF } X \text{ is } A_2 \text{ AND } Y \text{ is } B_2 \text{ THEN } Z \text{ is } C_2 \tag{2}$$

Here $X$ and $Y$ represent two antecedent variables (the conditional part of the production rule, like TP and WS as explained above), and $Z$ is the variable yielding the consequent part of the production rule. $A_1$, $A_2$, $B_1$, $B_2$, $C_1$, and $C_2$ are the linguistic and vague expressions with ambiguities. Focusing this idea of production rule, an example for such production rule that can be employed in the present research is shown as:

IF WP is *very high* AND TP is *lower* THEN RF is *moderate* (3)

Eq. (3) shows the qualitative form of explanation, such as *very high*, *lower* and *moderate* which are all fuzzy in nature. These are explained linguistically without specific quantity or as a crisp value. The relationship of the variables between antecedent and consequent parts represent a production rule in Eq. (3) based on valid logic. In the complex reality of the world, it is usually not easy to construct rules due to the limitations of manipulation and verbalization of experts (Abe and Ming-Shong, 1995). This method is termed as the fuzzy adaptive system (FAS).

## 2. Definitions

### 2.1. *Fuzzy set*

Fuzzy sets are the collection of objects with the same properties, and in crisp sets the objects either belong to the set or do not. In practice, the characteristic value for an object belonging to the considered set is coded as 1 and if it is outside the set then the coding is 0. In crisp sets, there is no ambiguity or vagueness about each object belongs to the considered set. On the other hand, in daily life humans are always confronted with objects that may be similar to one other with quite different properties. Therefore, uncertainty always arises concerning the assessment of membership values 0 or 1. Logically, of course, some of the similar objects may partially belong to the same set, therefore, an ambiguity emerges in the decision of belonging or not. In order to alleviate such situations Zadeh (1965) generalized the crisp set membership degree as having any value continuously between 0 and 1. Fuzzy sets are a generalization of conventional set theory. The basic idea of fuzzy sets is easy to grasp. An object with membership function 1 belongs to the set with no doubt and those with 0 membership functions again absolutely do not belong to the set, but objects with intermediate membership functions partially belong to the same set. The greater the membership function, the more the object belongs to the set (Hasan and Zenkai, 1999).

The membership function of a fuzzy set is a generalization of the indicator function in classical sets. In fuzzy logic, it represents the degree of truth as an extension of valuation. Degrees of truth are often confused with probabilities, although they are conceptually distinct, because fuzzy truth represents membership in vaguely defined sets, not likelihood of some event or condition.

For the universe $X$ and given the membership-degree function $\mu \rightarrow [0, 1]$ the fuzzy set is defined as

$$A = \{(x, \mu_A(x)) | x \in X\} \tag{4}$$

The following holds good for the functional values of the membership function $\mu_A(x)$:

$$\mu_A(x) \geq 0 \quad \forall x \in X \tag{5}$$

$$\overset{Sup}{x \in X}[\mu_A(x)] = 1 \tag{6}$$

### 2.2. Fuzzy levels

Range between the minimum (Min) and maximum (Max) value of any fuzzy variable is divided into suitable numbers which are denoted in ascending order starting from the minimum (Min) to maximum (Max) value of a fuzzy set. Fig. 1 shows the range and the fuzzy levels for a fuzzy set of objects, in a triangular functional diagram. Here the range has been divided into five fuzzy levels which are NL, NS, ZE, PS, and PL. A fuzzy inference model consists of three modules. Fig. 2 shows a schematic diagram of steps involved in fuzzy rule-based system. Definitions and methods of calculations are presented below.

### 2.3. Fuzzification

As per Lee (1990), fuzzification is a process which involves the following:

(1) measures the values of input variables,
(2) performs a scale mapping that transfers the range of values of input variables into a corresponding universe of discourse, and
(3) performs the function that converts input data into suitable linguistic values which may be viewed as labels of fuzzy sets.

Fig. 1 shows a value of a fuzzy variable $x$ intersecting the triangles with fuzzy levels of ZE and NS and their respective membership functions ($\mu$) of 0.3 and 0.7. Hence, fuzzification is the process that involves

(1) inputting the value of the fuzzy variable in the universe of discourse,

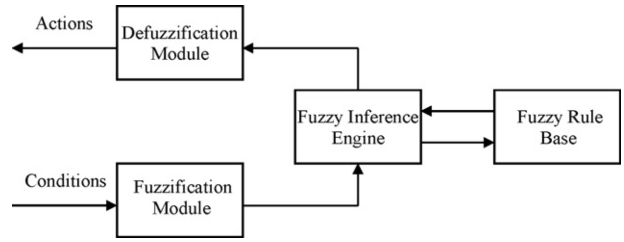



**Fig. 2 – General scheme of a fuzzy system.**

(2) obtaining the intersecting points on the arms of the triangles to calculate the fuzzy levels, and
(3) obtaining the corresponding membership functions ($\mu$).

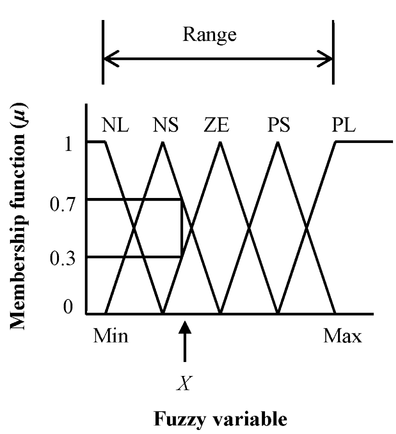
### 2.4. Min–Max composition

From Fig. 1, it is observed that one fuzzy variable ($x$) yields two membership functions (0.3 and 0.7) and their respective fuzzy levels are NS and ZE. Hence, if there are two fuzzy variables in the antecedent part, an increase in wind speed and a decrease in temperature when compared between the $i$th and $(i-1)$th day, hereafter denoted as WS and TP, respectively, as in Eqs. (7) and (8) below, there will be four membership functions and four respective fuzzy levels obtained after fuzzification. The mathematical method followed by fuzzification is termed as "min–max composition". Considering WS and TP as the two fuzzy variable inputs and rainfall, hereafter denoted as RF as the output in each of the production rules:

IF WS is *strong* and TP is *lower* THEN RF is *moderate* (7)

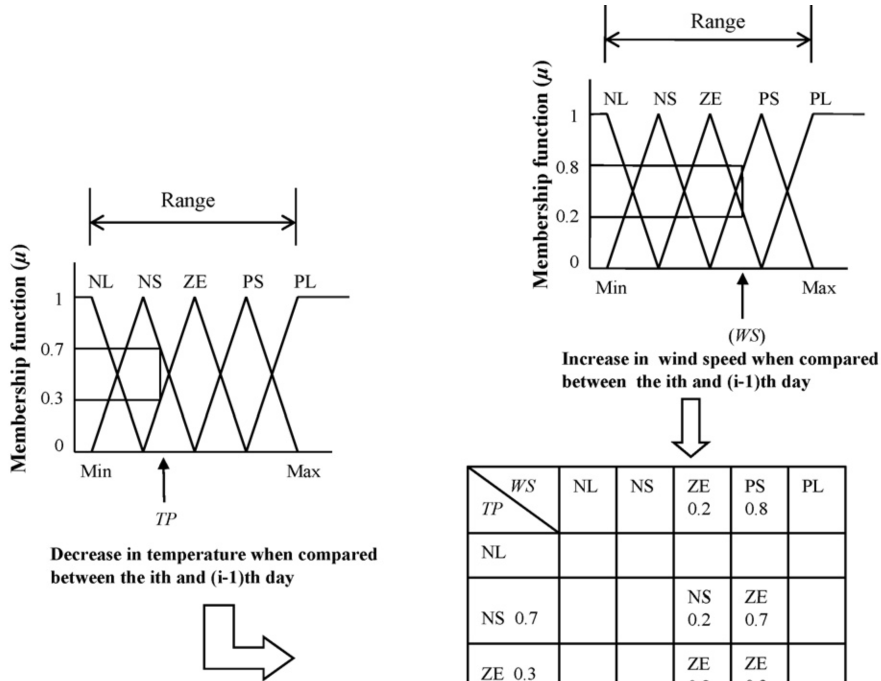
IF *WS* is *strong* and *TP* is *moderate* THEN *RF* is *moderate* (8)

Fuzzification of any of these production rules will yield fuzzy levels and membership functions as shown in Fig. 3. Here the values of WS and TP are the two fuzzy variables representing the antecedent part of a production rule yielding RF as its consequence that is shown in Fig. 3. Suppose a value of WS yields the membership function values of 0.2 and 0.8 belonging to the fuzzy levels of ZE and PS, respectively. Similarly, TP, another fuzzy variable in the antecedent part, yields membership functions values of 0.3 and 0.7 for the fuzzy levels of ZE




| Abbreviation | Stands for | Meaning |
|---|---|---|
| NL | Negative large | Very low |
| NS | Negative small | Low |
| ZE | Zero | Normal |
| PS | Positive small | High |
| PL | Positive large | Very high |

**Fig. 1 – Triangular functional diagram and method for calculating membership function ($\mu$) and corresponding fuzzy levels.**




| WS / TP | NL | NS | ZE 0.2 | PS 0.8 | PL |
|---|---|---|---|---|---|
| NL | | | | | |
| NS 0.7 | | | NS 0.2 | ZE 0.7 | |
| ZE 0.3 | | | ZE 0.2 | ZE 0.3 | |
| PS | | | | | |
| PL | | | | | |

**Fig. 3 – Method for calculating membership function ($\mu$) and corresponding fuzzy levels by min–max composition.**

and NS, respectively. Inferring fuzzy level for RF is NS which is shown in the production rule table of Fig. 3. The following equation holds good:

$$\text{IF WS is ZE and TP is NS THEN RF is NS} \tag{9}$$

Fig. 3 shows that a value of membership function $\mu$ for WS equals 0.2 with its fuzzy level of ZE. This figure further shows another value of membership function $\mu$ for TP equals 0.7 with its fuzzy level of NS. A value of membership function for RF is taken to be 0.2 as it is the minimum value of $\mu$ between 0.2 and 0.7. A similar mathematical approach for the same fuzzy variables of ZE for three other production rules inside the table is presented for RF in Fig. 3. Hence, the three minimum values 0.7, 0.2, and 0.3 for the same fuzzy levels of ZE are obtained. Finally, the maximum value 0.7 is taken out of the three minimum values of 0.7, 0.2, and 0.3 for the next step of the calculation process for defuzzification. In an example to show the generalized form of the equation for min–max composition, two equations for the four production rules presented in the table written here as follows:

$$\text{IF WS is } LW_1 \text{ and TP is } LT_1 \text{ then RF is ZE} \tag{10}$$

$$\text{IF WS is } LW_2 \text{ and TP is } LT_2 \text{ then RF is ZE} \tag{11}$$

Eqs. (10) and (11) have the same fuzzy levels of ZE for RF. Hence, the general form of the equation for calculating the membership function $\mu^{(ZE)}(RF)$ having the same fuzzy levels ZE for the consequent part can be shown as

$$\mu^{(ZE)}(RF) = \bigcup_{i=1}^{3}\left[\mu^{(LW_i)}(WS) \cap \mu^{(TP_i)}(LT)\right] \tag{12}$$

Here, $\mu^{(ZE)}(RF)$ is the membership function for RF for fuzzy level ZE, LW is the fuzzy level for WS, LT is the fuzzy level for TP, and $\cap$ indicates selecting the minimum value of membership function out of $\mu^{(LW_i)}(WS)$ and $\mu^{(TP_i)}(LT)$. $\cup$ indicates selecting the maximum value of the calculated minimum membership function values. $i$ is the number of production rules having the same fuzzy levels (here it is ZE). Eq. (12) is valid only when $i > 1$.

If the fuzzy levels of RF are not the same, then the membership functions of RF can be calculated by the following equation:

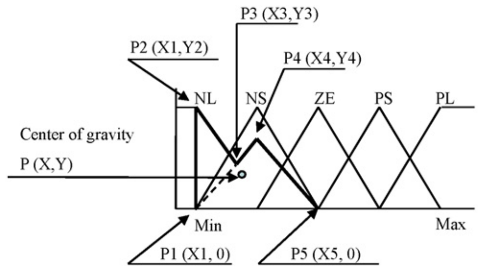
$$\mu^{(LV)}(RF) = \mu^{(LW_i)}(WS) \cap \mu^{(TP_i)}(LT) \tag{13}$$

Here, LV, the abbreviation for fuzzy level for RF, is different for various production rules. In these cases, only the minimum value of the membership functions between $\mu^{(LW_i)}(WS)$ and $\mu^{(TP_i)}(LT)$ is considered.

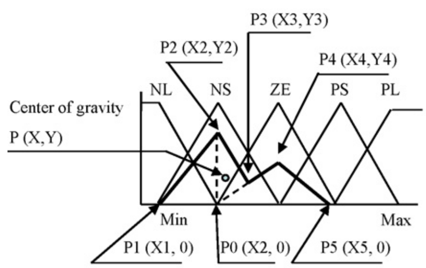
### 2.5. *Defuzzification*

Defuzzification is the calculation method to yield the quantified value for the consequent part of a fuzzy statement described by production rule. Defuzzification performs the following functions:

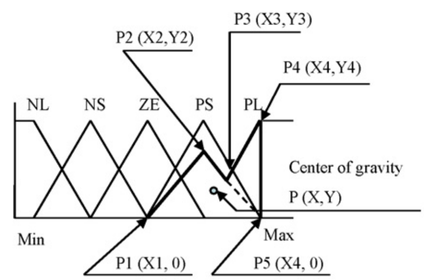
(1) a scale mapping which converts the range of values of output variables into corresponding universe of discourse and




(a) **Fuzzy levels for rainfall between NL and NS.**




(b) **Fuzzy levels for rainfall between NS and PS.**




(c) **Fuzzy levels for rainfall between PS and PL.**

**Fig. 4 – Calculation methods for defuzzification. (a) Fuzzy levels for rainfall between NL and NS. (b) Fuzzy levels for rainfall between NS and PS. (c) Fuzzy levels for rainfall between PS and PL.**

(2) yields a non-fuzzy control action from an inferred fuzzy control action.

Fig. 4 illustrates the mathematical procedure followed to calculate the center of gravity for the defuzzification method. The following are the possible cases:

Case1. Fuzzy levels of the inference part of production rules belong to NL and NS with their corresponding values of membership functions ($\mu$).
Case 2. Fuzzy levels of the inference part of production rules are in the region from NS to PS range with their corresponding values of membership functions ($\mu$).
Case 3. Fuzzy levels of the inference part of production rules belong to PS and PL with their corresponding values of membership functions ($\mu$).

Considering Fig. 4(a) as a description of the mathematical procedure for calculating center of gravity for Case 1, the point of intersections may be defined as P1(X1, 0), P2(X1, Y2), P3(X3, Y3), P4(X4, Y4) and P5(X5, 0). Let the co-ordinate of center of gravity for the area bounded by the above five co-ordinates be P(X, Y). There are two triangles, one which can be shown by the co-ordinates P1(X1, 0), P2(X1, Y2) and P3(X3, Y3); and the other triangle can be shown by the co-ordinates P1(X1, 0), P4(X4, Y4) and P5(X5, 0).

Now, the average of X values in triangle formed by P1(X1, 0), P2(X1, Y2), and P3(X3, Y3) is

$$\mathrm{XC1} = \frac{\mathrm{X1} + \mathrm{X1} + \mathrm{X3}}{3.0} \tag{14}$$

and the Y value in the same triangle formed by P1(X1, 0), P2(X1, Y2), and P3(X3, Y3) is

$$\mathrm{YC1} = \frac{0 + \mathrm{Y2} + \mathrm{Y3}}{3.0} \tag{15}$$

Similarly, X value in triangle formed by P1(X1, 0), P4(X4, Y4) and P5(X5, 0) is

$$\mathrm{XC2} = \frac{\mathrm{X1} + \mathrm{X4} + \mathrm{X5}}{3.0} \tag{16}$$

and the Y value in the same triangle formed by P1(X1, 0), P4(X4, Y4) and P5(X5, 0) is

$$\mathrm{YC2} = \frac{0 + \mathrm{Y4} + 0}{3.0} \tag{17}$$

Area formed by P1(X1, 0), P2(X1, Y2), and P3(X3, Y3) is

$$\text{area 1} = \frac{(\mathrm{X1} - \mathrm{X1})(\mathrm{Y3} - 0) - (\mathrm{X3} - \mathrm{X1})(\mathrm{Y2} - 0)}{2.0} \tag{18}$$

Similarly, area formed by P1(X1, 0), P4(X4, Y4) and P5(X5, 0) is

$$\text{area 2} = \frac{(\mathrm{X4} - \mathrm{X1})(0 - 0) - (\mathrm{X5} - \mathrm{X1})(\mathrm{Y4} - 0)}{2.0} \tag{19}$$

Therefore,

$$\text{total area} = \text{area 1} + \text{area 2} \tag{20}$$

and the area covered by P1(X1, 0), P2(X1, Y2), and P3(X3, Y3) is

$$\text{Fraction 1} = \frac{\text{area 1}}{\text{area 1} + \text{area 2}} = \frac{\text{area 1}}{\text{total area}} \tag{21}$$

and the area covered by P1(X1, 0), P4(X4, Y4) and P5(X5, 0) is

$$\text{Fraction 2} = \frac{\text{area 2}}{\text{area 1} + \text{area 2}} = \frac{\text{area 2}}{\text{total area}} \tag{22}$$

Therefore, the co-ordinate for center of gravity is

$$X = (XC1 \times \text{Fraction 1} + XC2 \times \text{Fraction 2}) \tag{23}$$

and

$$Y = (YC1 \times \text{Fraction 1} + YC2 \times \text{Fraction 2}) \tag{24}$$

Considering Fig. 4(b) as describing the mathematical procedure for calculating the center of gravity for Case 2, the point of intersections may be defined as P0(X2, Y0), P1(X1, 0), P2(X2, Y2), P3(X3, Y3), P4(X4, Y4) and P5(X5, 0). Let the co-ordinate of the center of gravity of the area bounded by the above five co-ordinates represented by thick lines be P(X, Y). Let us consider that total area under the thick lines consists of three small triangles which are as follows:

triangle 1 which is formed by the co-ordinates (P0, P1, and P2),
triangle 2 which is formed by the co-ordinates (P0, P2, and P3), and
triangle 3 which is formed by the co-ordinates (P0, P4, and P5).

The co-ordinates for triangle 1 are P0(X2, 0), P1(X1, 0), and P2(X2, Y2); triangle 2 consisting of P0(X2, 0), P2(X2, Y2) and P3(X3, Y3); and triangle 3 shown with the co-ordinates P0(X2, 0), P4(X4, Y4) and P5(X5, 0).

Now, the average of X value in triangle formed by P0(X2, 0), P1(X1, 0), and P2(X2, Y2) is

$$XC1 = \frac{X2 + X1 + X2}{3.0} \tag{25}$$

and the Y value in the same triangle formed by P0(X2, 0), P1(X1, 0), and P2(X2, Y2) is

$$YC1 = \frac{0 + 0 + Y2}{3.0} \tag{26}$$

Similarly, X value in triangle formed by P0(X2, 0), P2(X2, Y2) and P3(X3, Y3) is

$$XC2 = \frac{X2 + X2 + X3}{3.0} \tag{27}$$

and the Y value in the same triangle formed by P0(X2, 0), P2(X2, Y2) and P3(X3, Y3) is

$$YC2 = \frac{0 + Y2 + Y3}{3.0} \tag{28}$$

Similarly, X value in triangle formed by P0(X2, 0), P4(X4, Y4) and P5(X5, 0) is

$$XC2 = \frac{X2 + X4 + X5}{3.0} \tag{29}$$

and the Y value in the same triangle formed by P0(X2, 0), P4(X4, Y4) and P5(X5, 0) is

$$YC2 = \frac{0 + Y4 + 0}{3.0} \tag{30}$$

Area formed by P0(X2, 0), P1(X1, 0), and P2(X2, Y2) is

$$\text{area 1} = \frac{(X1 - X2)(Y2 - 0) - (X2 - X2)(0 - 0)}{2.0} \tag{31}$$

Similarly, area formed by P0(X2, 0), P2(X2, Y2) and P3(X3, Y3) is

$$\text{area 2} = \frac{(X2 - X2)(Y3 - 0) - (X3 - X2)(Y2 - 0)}{2.0} \tag{32}$$

Similarly, area formed by P0(X2, 0), P4(X4, Y4) and P5(X5, 0) is

$$\text{area 3} = \frac{(X4 - X2)(0 - 0) - (X5 - X2)(Y4 - 0)}{2.0} \tag{33}$$

Hence,

$$\text{total area} = \text{area 1} + \text{area 2} + \text{area 3} \tag{34}$$

area covered by P0(X2, 0), P1(X1, 0), and P2(X2, Y2) is

$$\text{Fraction 1} = \frac{\text{area 1}}{\text{total area}} \tag{35}$$

area covered by P0(X2, 0), P2(X2, Y2) and P3(X3, Y3) is

$$\text{Fraction 2} = \frac{\text{area 2}}{\text{total area}} \tag{36}$$

and area covered by P0(X2, 0), P4(X4, Y4) and P5(X5, 0) is

$$\text{Fraction 3} = \frac{\text{area 3}}{\text{total area}} \tag{37}$$

Therefore, the co-ordinate for center of gravity is

$$X = XC1 \times \text{Fraction 1} + XC2 \times \text{Fraction 2} + XC3 \times \text{Fraction 3} \tag{38}$$

and

$$Y = YC1 \times \text{Fraction 1} + YC2 \times \text{Fraction 2} + YC3 \times \text{Fraction 3} \tag{39}$$

Considering Fig. 4(c) for describing the mathematical procedure for calculating the center of gravity for Case 3, the point of intersections may be defined as P1(X1, 0), P2(X2, Y2), P3(X3, Y3), P4(X4, Y4) and P5(X4, 0). Let the co-ordinate of the center of gravity of the area bounded by the above five co-ordinates represented by thick lines be $P(X, Y)$. There are two triangles one of which can be shown by the co-ordinates P1(X1, 0), P2(X2, Y2) and P5(X4, 0) and the other triangle can be shown with the co-ordinates P3(X3, Y3), P4(X4, Y4) and P5(X4, 0).

Now, the average of $X$ value in triangle formed by P1(X1, 0), P2(X2, Y2) and P5(X4, 0) is

$$XC1 = \frac{X1 + X2 + X4}{3.0} \quad (40)$$

and the Y value in the same triangle formed by P1(X1, 0), P2(X2, Y2) and P5(X4, 0) is

$$YC1 = \frac{0 + Y2 + 0}{3.0} \quad (41)$$

Similarly, $X$ value in triangle formed by P3(X3, Y3), P4(X4, Y4) and P5(X4, 0) is

$$XC2 = \frac{X3 + X4 + X4}{3.0} \quad (42)$$

and the $Y$ value in the same triangle formed by P3(X3, Y3), P4(X4, Y4) and P5(X4, 0) is

$$YC2 = \frac{Y3 + Y4 + 0}{3.0} \quad (43)$$

Area formed by P1(X1, 0), P2(X2, Y2) and P5(X4, 0) is

$$\text{area 1} = \frac{(X1 - X4)(Y2 - 0) - (X2 - X4)(0 - 0)}{2.0} \quad (44)$$

Similarly, area formed by P3(X3, Y3), P4(X4, Y4) and P5(X4, 0) is

$$\text{area 2} = \frac{(X3 - X4)(Y4 - 0) - (X4 - X4)(Y3 - 0)}{2.0} \quad (45)$$

Therefore,

$$\text{total area} = \text{area 1} + \text{area 2} \quad (46)$$

and the area covered by P1(X1, 0), P2(X2, Y2) and P5(X4, 0) is

$$\text{Fraction 1} = \frac{\text{area 1}}{\text{area 1} + \text{area 2}} = \frac{\text{area 1}}{\text{total area}} \quad (47)$$

and the area covered by (X3, Y3), P4(X4, Y4) and P5(X4, 0) is

$$\text{Fraction 2} = \frac{\text{area 2}}{\text{area 1} + \text{area 2}} = \frac{\text{area 2}}{\text{total area}} \quad (48)$$

Therefore, the co-ordinate for center of gravity is

$$X = XC1 \times \text{Fraction 1} + XC2 \times \text{Fraction 2} \quad (49)$$

and

$$Y = YC1 \times \text{Fraction 1} + YC2 \times \text{Fraction 2} \quad (50)$$

## 3. Study area

This manuscript presents a fuzzy inference model for predicting RF using meteorological scan data from the United States Department of Agriculture (USDA) Soil Climate Analysis Network Station at Alabama Agricultural and Mechanical University (AAMU) campus. Meteorological data for 2003 and 2004 were collected and analyzed to determine the variables that are involved in rainfall occurrences. The Alabama Mesonet (ALMNet) has been the apex representing 14 combinations of meteorological/soil profile stations and 12 soil profile stations distributed in 11 counties in southern Tennessee and north and central Alabama. The combination stations are also part of the USDA and Natural Resources Conservation Service (NRCS) scan network. Alabama Mesonet (ALMNet) is controlled and run by the Environment, Soil and Water Science Program, department of NRES of Alabama Agricultural and Mechanical University (AAMU).

## 4. Model development

Although meteorological scan data were collected for 2 years, 2003 and 2004, the model was developed based on year 2004 data. These data were very well organized including soil related parameters. Data for Bragg Farm and Winfred A. Thomas Agricultural Research Station (WTARS) were also collected, monthly data spread sheets were prepared, and graphs plotted to assist with pre-assessment of analysis and to generate ideas on climatic behavior. Fig. 5 shows the characteristics of rainfall for the month of August 2004 using data from the AAMU campus. Based on the observations of the graphs prepared for every month during the years 2003 and 2004 for AAMU, Bragg and WTARS farms, it was apparent that a value of WS and another value of TP when compared between the ith and $(i-1)$th day mostly resulted in a rainfall

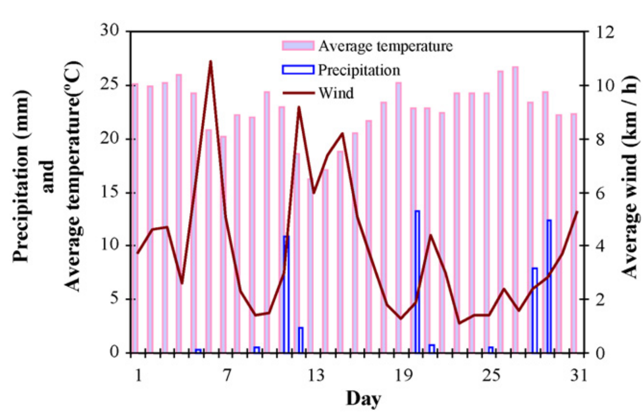
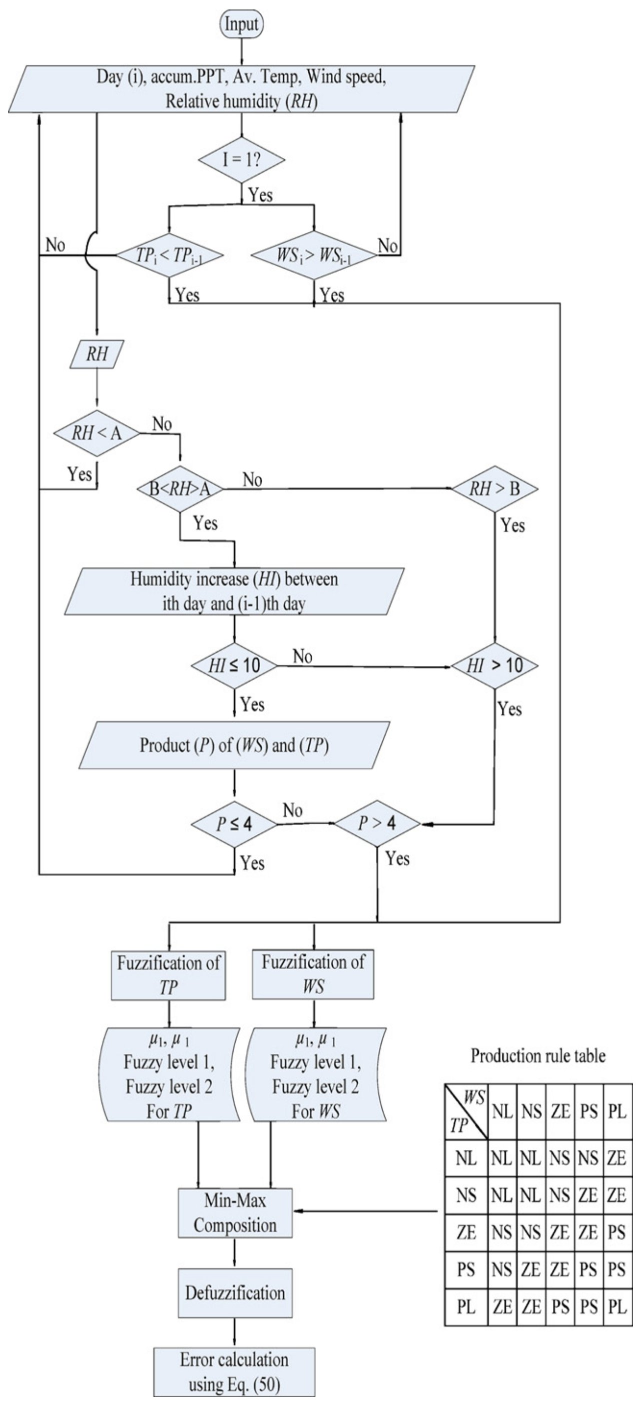



**Fig. 5 – Rainfall pattern against wind and temperature using USDA scan data from AAMU campus for August 2004.**

occurrence. Usually, the characteristic of rainfall occurrence takes place at the first or second day of the phenomena of increasing of wind speed and decreasing of temperature. Hence, the degree of association between WS and TP when compared between the ith and $(i-1)$th day causing RF occurrences was established. Based on analysis, it was observed that a RF occurrence has a positive relation with TP and WS. The observation further revealed that the relation of RF occurrence with TP and WS reflects expert knowledge. Hence, the values of WS and TP between ith and $(i-1)$th day and using them in the fuzzy inference model for the antecedent part of production rules was considered to be feasible. Fig. 6 shows the fuzzy inference model structure and the steps followed to determine the time and amount of RF. This figure has been prepared by incorporating the consideration of threshold values as described in Fig. 7. In the initial step of calculation, temperature, wind speed, and relative humidity were converted to yield the average daily values dividing by 24 (1 day = 24 h) to produce average temperature, average wind speed and average relative humidity.




| TP \ WS | NL | NS | ZE | PS | PL |
|---|---|---|---|---|---|
| NL | NL | NL | NS | NS | ZE |
| NS | NL | NL | NS | ZE | ZE |
| ZE | NS | NS | ZE | ZE | PS |
| PS | NS | ZE | ZE | PS | PS |
| PL | ZE | ZE | PS | PS | PL |

**Fig. 6 – Model structure and steps followed for timing and amount of rainfall (RF).**

A preliminary analysis showed that the variables described below had a significant influence over RF occurrences:

(a) relative humidity (RH) of the ith day,
(b) humidity increase (HI) is increase in relative humidity when compared between the ith and $(i-1)$th day, and
(c) product ($P$) of TP and WS.

These three variables were taken into consideration and shown in Fig. 6 in the calculation process with seasonal variation as shown in tabular form in Fig. 7. This variation was considered with two threshold values for minimum and maximum limits as indicated by A and B in Fig. 7:

(a) 1 January–30 April,
(b) 1 May–30 September, and
(c) 1 October–31 December

The threshold values were selected based on the calculation of results of the model.

In year 2004, there were 6 days out of 132 total rainy days when the actual amount of RF was more than 50 mm. Considering the uniformity of data range, and to avoid very unusual phenomena, the highest volume of RF was considered to be 50 mm for the maximum value of predicted RF for defuzzification process (refer Fig. 4). The error was calculated using the following equation:

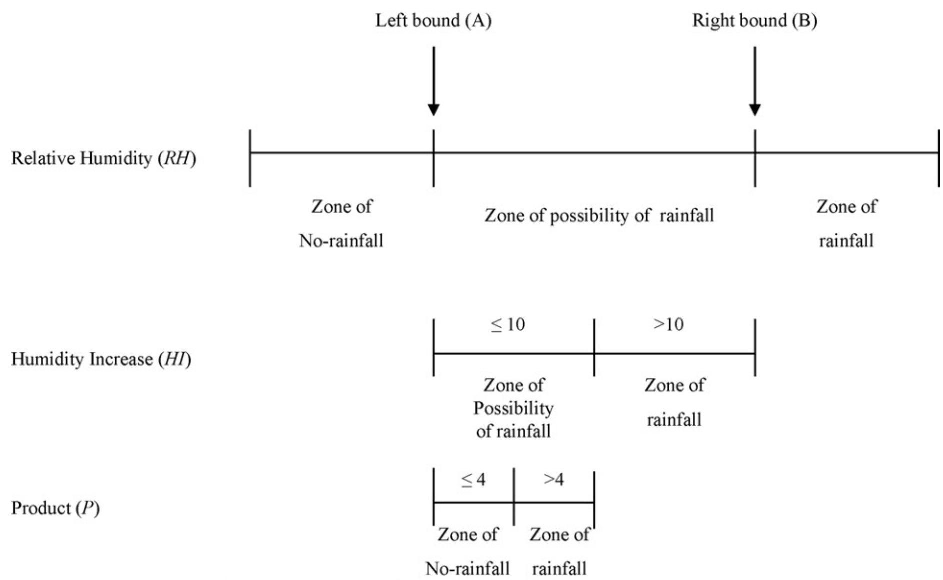
$$\frac{1}{n}\sum_{i=1}^{n}\mathrm{abs}\left[\frac{\mathrm{RF}_{a_i}-\mathrm{RF}_{c_i}}{\mathrm{RF}_{a_i}}\right]\times 100 \qquad (51)$$

Here, $n$ is the number of days of rainfall occurrences, $\mathrm{RF}_{a_i}$ is actual amount of rainfall, and $\mathrm{RF}_{c_i}$ is the calculated amount of rainfall.

## 5. Results and discussion

### 5.1. *Selection of variables*

Fuzzy variables of WS and TP between the ith and $(i-1)$th day were good choices for the development of a model for predicting RF. In reality, fuzzy inference models involve variables which are perceived by experts as responsible for the consequence part of the production rule. This means a fuzzy inference model reflects the scenario of thinking and decision-making process by expert knowledge. The fuzzy variables were chosen following the assessment on graphs prepared on the basis of monthly data from AAMU for 2003




| | (A) | (B) |
|---|---|---|
| Jan 1 to Apr 30 | 70 | 85 |
| May 1 to Sep 30 | 80 | 90 |
| Oct 1 to Dec 31 | 75 | 90 |

**Fig. 7 – Assumptions and threshold values for determining rainfall and no-rainfall occurrences.**

and 2004. Selection of variables of TP and WS between the ith and (i − 1)th day was considered for this model as a better approach. The final results indicated that the selection of these two variables was suitable for the development of the model and that they showed a good agreement when used in the antecedent part of the production rule.

### 5.2. *Selection of fuzzy levels for the inference part of the production rule table*

Selection of the fuzzy levels in the inference part of the production rules is a cumbersome process by trial and error method. Twenty-five (5 × 5) fuzzy variables for the inference part are shown in the table in Fig. 3. A method for iterating the fuzzy variables for RF was followed in the computer program that selected the one yielding the lowest percentage of error based on Eq. (51). Depending on the scenario of the system, fuzzy levels in the inference part of the production rule must have either an ascending or a descending nature. Skill and logical approachability are required for determining fuzzy variables for the consequent part with respect to the fuzzy variables in the antecedent part of the production rule. The production rule table shown in Fig. 6 was the best set of fuzzy levels for RF that yielded the lowest error value of 12.35%.

### 5.3. *Maximum value of RF*

Real RF data showed that the maximum RF was 93 mm which is very unusual and rare for the same location. Moreover, if the actual amount of RF is considered to be more than 50 mm, the region of maximum RF (around PL of Fig. 4(c)) will have unrealistic and lesser density of number of data compared to density of data in the region of NL, NS, ZE, and PS. Hence, considering the uniformity of data distribution among the ranges of NL, NS, ZE, PS, and PL the maximum value of predicted RF to be 50 mm was justifiable.

### 5.4. *Selection of threshold values for predicted value of RF*

Based on the fundamental logic of this research that a value of WS and another value of TP when compared between the ith and (i − 1)th day may result in RF, their fuzzy levels, production rules, and ranges of variables, showed dependency on three other possible factors. These factors need to be considered with their threshold values for matching the actual and calculated amount of RF. These factors are (1) average daily relative humidity, (2) humidity increase between the ith and (i − 1)th day, and (3) product (*P*) of TP and WS between the ith and (*i* − 1)th day. Fig. 7 represents two boundary values (A) and (B) for RH. The zone between (A) and (B) is the range for a possible RF and the zone beyond (B) is the zone for RF regardless of any other consideration, whereas the RH of less than (A) is the zone for no RF. When the value of HI is more than 10 and it is within the boundary values of (A) and (B) then it becomes the zone for RF. The zone for the value of HI of less than 10 is again the zone for a possible RF occurrence. This possibility is further considered to occur when the value of product (*P*) of TP and WS is greater than 4. But if the value of *P*

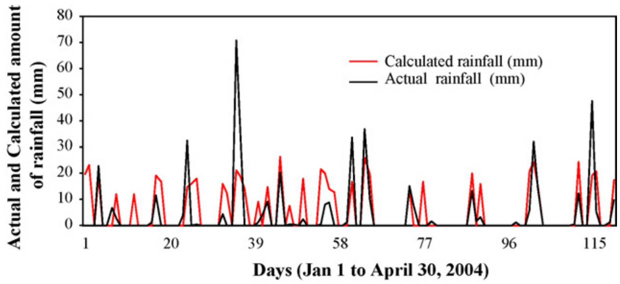



**Fig. 8 – Actual and calculated amount of rainfall using USDA scan data from AAMU (1 January–30 April, 2004).**

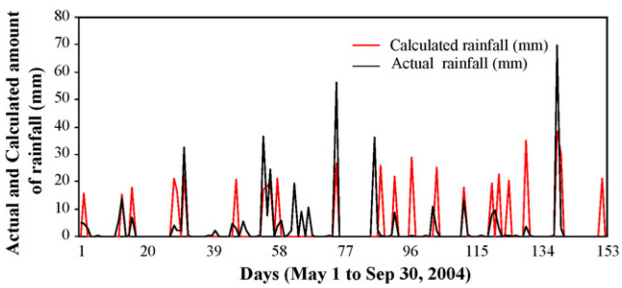



**Fig. 9 – Actual and calculated amount of rainfall using USDA scan data from AAMU (1 May–30 September, 2004).**

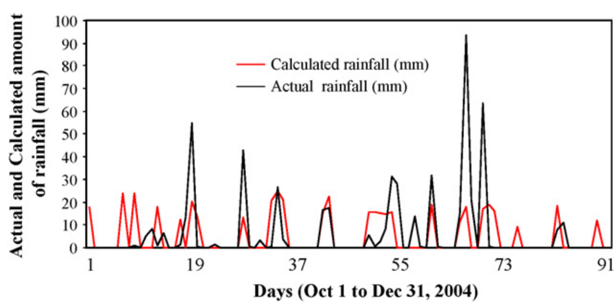



**Fig. 10 – Actual and calculated amount of rainfall using USDA scan data from AAMU (1 October–31 December, 2004).**

of TP and WS is less than 4, then it was considered that there would be no RF occurrence. Introducing these three threshold values with consideration of three different seasons as described in Fig. 7, the model showed good agreement between the actual amount of RF and predicted value for RF. Figs. 8–10 show the actual and predicted values of RF using 2004 scan data from the USDA Soil Climate Analysis Network Station at the AAMU campus. These figures illustrate the actual amount of RF and predicted value of RF during three different seasons as considered in this model and explained in Fig. 7. The figures further show that the timeliness of the actual amount of RF and predicted value of RF almost perfectly match, but the amount of RF needs further research to yield better agreement between actual and predicted values of RF. Therefore, further research planned to develop an approach for auto-generation of the production rules by iteration method and selecting the particular production rule table that yields the lowest percentage of error.

## 6. Conclusion

Selection of variables and the fundamental logic that the values TP and WS was an attempt to identify amount of RF and its time of occurrence as the consequent part of the fuzzy inference model. Introducing the idea of threshold values of (a) RH of the $i$th day, (b) HI when compared between the $i$th and $(i-1)$th day, and (c) $P$, product of WS and TP appeared to be an appropriate attempt for the model to match the actual RF occurrences. Iteration of the fuzzy levels with logic both for antecedent and consequent parts was found to be efficient. Further research has been planned to attain the maximum possible matches of time and amount of RF between actual occurrences and the one predicted by the model.

## Acknowledgements

The authors appreciate the faculty and staff in the Department of Natural Resources and Environmental Sciences (NRES) of Alabama Agricultural and Mechanical University (AAMU) for their constructive suggestions and support. Thanks to Mr. Daryl Lawson, Prof. Rao Mentreddy of Alabama Agricultural and Mechanical University and Dr. Rafiq Islam of Ohio State University for reviewing this manuscript. United States Department of Agriculture (USDA) and Natural Resources Conservation Service (NRCS) are duly acknowledged for providing fund for conducting this research. The authors also acknowledge the late Robert Metzl for providing data related to this research.

## REFERENCES

Abe, S., Ming-Shong, L., 1995. Fuzzy rules extraction directly from numerical data for function approximation. IEEE Transactions on System, Man and Cybernetics 25 (1), 119–129.

Brown-Brandl, T.M., Jones, D.D., Woldt, W.E., 2003. Evaluating modeling techniques for livestock heat stress prediction. In: 2003 ASAE Annual Meeting. Paper Number 034009. http://www.frymulti.com/abstract.asp?aid=14084&t=2.

Carrano, L.A., Taylor, B.J., Robert, E.Y., Richard, L.L., Daniel, E.S., 2004. Fuzzy knowledge-based modeling and regression in abrasive wood machining. Forest Products Journal 54 (5), 66–72.

Hasan, M., Mizutani, M., Goto, A., Matsui, H., 1995. A model for determination of intake flow size—development of optimum operational method for irrigation using fuzzy set theory (1). System Nogaku. Journal of the Japanese Agricultural System Society 11 (1), 1–13.

Hasan, T., Zenkai, S., 1999. A new modeling approach for predicting the maximum daily temperature from a time series. Turkish Journal of the Engineering and Environmental Science 23, 173–180.

Lee, C.C., 1990. Fuzzy logic in control systems: fuzzy logic controller. Part I. IEEE Transaction of System, Man and Cybernetics 20 (March/April (2)), 404–418.

Lee, S., Cho, S., Wong, P.M., 1998. Rainfall prediction using artificial neural networks. Journal of Geographic Information and Decision Analysis 2 (2), 233–242.

Pant, L.M., Ashwagosh, G., 2004. Fuzzy rule-based system for prediction of direct action avalanches. Current Science 87 (1), 99–104.

Wilks, D.S., 1998. Multisite generalization of a daily stochastic precipitation generating models. Journal of Hydrology 210, 178–191.

Wong, K.W., Wong, P.M., Gedeon, T.D., Fung, C.C., 2003. Rainfall prediction model using soft computing technique. Soft Computing—A Fusion of Foundations, Methodologies and Applications, vol. 7, No. 6. Springer, Berlin/Heidelberg. ISBN 1432-7643 (Print), 1433-7479 (Online).

Zadeh, A.L., 1965. Fuzzy logic. Information and Control 8 (3), 338–353.